

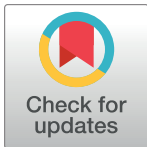
RESEARCH ARTICLE

# Plasmodium apicoplast tyrosyl-tRNA synthetase recognizes an unusual, simplified identity set in cognate tRNA<sup>Tyr</sup>

Marta Cela<sup>1</sup>, Caroline Paulus<sup>1</sup>, Manuel A. S. Santos<sup>2</sup>, Gabriela R. Moura<sup>2</sup>, Magali Frugier<sup>1\*</sup>, Joëlle Rudinger-Thirion<sup>1</sup>

**1** UPR 9002 Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg Cedex, France, **2** Department of Medical Sciences and Institute of Biomedicine - iBiMED, University of Aveiro, Aveiro, Portugal

\* [m.frugier@ibmc-cnrs.unistra.fr](mailto:m.frugier@ibmc-cnrs.unistra.fr)



## OPEN ACCESS

**Citation:** Cela M, Paulus C, Santos MAS, Moura GR, Frugier M, Rudinger-Thirion J (2018) *Plasmodium* apicoplast tyrosyl-tRNA synthetase recognizes an unusual, simplified identity set in cognate tRNA<sup>Tyr</sup>. PLoS ONE 13(12): e0209805. <https://doi.org/10.1371/journal.pone.0209805>

**Editor:** Andrew M. Blagborough, Imperial College London, UNITED KINGDOM

**Received:** October 26, 2018

**Accepted:** December 11, 2018

**Published:** December 28, 2018

**Copyright:** © 2018 Cela et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This work was supported by CNRS (<http://www.cnrs.fr/>) to JRT, MF; the University of Strasbourg (<https://www.unistra.fr/>) to CP and the LABEX: ANR-10-LABX-0036\_NETRNA ([www.agence-nationale-recherche.fr/](http://www.agence-nationale-recherche.fr/)) and Fondation pour la Recherche Médicale [grant numbers FDT201704337050] ([www.frm.org/](http://www.frm.org/)) to MC. The funders had no role in study design, data collection

## Abstract

The life cycle of *Plasmodium falciparum*, the agent responsible for malaria, depends on both cytosolic and apicoplast translation fidelity. Apicoplast aminoacyl-tRNA synthetases (aaRS) are bacterial-like enzymes devoted to organellar tRNA aminoacylation. They are all encoded by the nuclear genome and are translocated into the apicoplast only after cytosolic biosynthesis. Apicoplast aaRSs contain numerous idiosyncratic sequence insertions: An understanding of the roles of these insertions has remained elusive and they hinder efforts to heterologously overexpress these proteins. Moreover, the A/T rich content of the *Plasmodium* genome leads to A/U rich apicoplast tRNA substrates that display structural plasticity. Here, we focus on the *P. falciparum* apicoplast tyrosyl-tRNA synthetase (*Pf*-apiTyrRS) and its cognate tRNA<sup>Tyr</sup> substrate (*Pf*-apitRNA<sup>Tyr</sup>). Cloning and expression strategies used to obtain an active and functional recombinant *Pf*-apiTyrRS are reported. Functional analyses established that only three weak identity elements in the apitRNA<sup>Tyr</sup> promote specific recognition by the cognate *Pf*-apiTyrRS and that positive identity elements usually found in the tRNA<sup>Tyr</sup> acceptor stem are excluded from this set. This finding brings to light an unusual behavior for a tRNA<sup>Tyr</sup> aminoacylation system and suggests that *Pf*-apiTyrRS uses primarily negative recognition elements to direct tyrosylation specificity.

## Introduction

*Plasmodium falciparum* causes the most severe form of malaria in humans. Rapid constitutive growth and expansion of the parasite are highly dependent on the continuous synthesis of proteins in the cytosol and organellar compartments [1]. Indeed, *Plasmodium* contains three genomes: nuclear, apicoplast (a relic chloroplast) and mitochondrial. These genomes require dedicated translation machineries to function, even if translation has not yet been explicitly demonstrated in the mitochondria. The *Plasmodium* apicoplast is essential and is involved in the synthesis of fatty acids, isoprenoid precursors and heme [2]. It has a 35 kb circular genome encoding 30 protein genes, all of which are involved in apicoplast transcription and translation

and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

including, for example, the subunits of RNA polymerase, elongation factor Tu, several ribosomal proteins, as well as ribosomal RNAs and transfer RNAs (tRNAs) [3]. However, most proteins essential for apicoplast functions are encoded by the nuclear genome and are imported into the organelle after translation. The *P. falciparum* genome contains more than 450 proteins genes with targeting signals for the apicoplast [4], among them are the apicoplast aminoacyl-tRNA synthetases (aaRSs), dedicated to the specific aminoacylation of apicoplast tRNAs with their corresponding amino acids.

Generally, there are 20 different aaRSs in each protein translation compartment, with each enzyme responsible for aminoacylating a specific tRNA isoacceptor set with the matching, cognate amino acid [5]. In *Plasmodium*, 37 nuclear genes encode 36 aaRSs (the cytosolic phenylalanyl-tRNA synthetase being a heterotetramer) [1,6]: 17 are exclusively cytosolic, 20 possess an apicoplast targeting sequence (4 of these, alanyl-, glycyl-, threonyl- and cysteinyl-tRNA synthetases, are targeted to both the apicoplast and the cytosol [7,8]) and only one is putatively mitochondrial [9].

Although all aaRSs perform the same reaction, namely the attachment of an amino acid to the 3' end of their cognate tRNA(s), they are structurally different. Depending on the architecture of the active site and the tRNA binding mode, aaRSs are divided into two structural classes, with 10 enzymes in each class [5,10,11]. In addition to the catalytic domain, they typically contain a tRNA-binding domain, which often recognizes the anticodon triplet of the tRNA. This organization generally holds true for organellar aaRSs, with some variations in size, oligomerization, and function [12]. Apicoplast aaRSs are evolutionarily conserved enzymes, yet previous studies have highlighted the presence of numerous uncharacterized insertions, which therefore increase their sizes significantly [6]. The presence of these insertions could explain why apicoplast aaRSs are poorly studied *in vitro*, despite their potential as therapeutic targets for the development of new anti-malarial drugs [13–15].

It is generally accepted that the fidelity of protein synthesis depends mostly on the formation of correct aminoacylation of tRNAs, the reaction catalyzed by the aaRSs [5,16]. Thus, a given aaRS selects one amino acid from the 20 canonical amino acids and one isoacceptor set of tRNAs from all tRNAs. The specificity of this process is mainly governed by identity elements scattered throughout tRNA structures. Positive identity elements are identified as nucleosides or structural motifs that are recognized in a specific manner by the cognate aaRS to target that tRNA for aminoacylation. Negative identity elements play an essential role in preventing the binding of non-cognate aaRSs to a given tRNA. tRNA identity sets have been characterized for most bacterial aminoacylation systems, but fewer examples have been studied in depth in archaeal, eukaryal or organellar systems [17,18]. Moreover, in view of the conserved, canonical cloverleaf structure of tRNAs, the nature of these identity elements are generally preserved across evolution with only a few exceptions. In particular, the tyrosine aminoacylation system is characterized by a robust phylogenetic barrier that limits cross-reactivity between partners from different species. For example, *E. coli* TyrRS does not aminoacylate eukaryotic tRNA<sup>Tyr</sup> and *vice-versa* [19].

Guided by the above background, we set out to characterize the *P. falciparum* apicoplast TyrRS/tRNA<sup>Tyr</sup>. We characterized the structural features required to specifically aminoacylate the apicoplast tRNA<sup>Tyr</sup> (*Pf*-apitRNA<sup>Tyr</sup>, encoded in the apicoplast genome) using the apicoplast tyrosyl-tRNA synthetase (*Pf*-apiTyrRS, encoded in the nuclear genome). These efforts required: (i) the engineering of the *Pf*-apiTyrRS gene to produce a soluble recombinant protein containing two *Plasmodium*-specific insertions; and (ii) confirmation that the A/U rich *Pf*-apitRNA<sup>Tyr</sup> transcript could be properly folded. Then, a systematic study of nucleotide replacements in tRNA, including a comparative examination of their activities, established the rules that govern recognition of *Pf*-apitRNA<sup>Tyr</sup> by its cognate TyrRS. Our analysis

demonstrates that the identity elements in *Pf*-apitRNA<sup>Tyr</sup> are unusually reduced in strength and number. These results reveal that the identity elements of the apicoplast tyrosine aminoacylation system are both distinct and minimalistic in comparison to those that have been conserved evolutionarily elsewhere.

## Material and methods

### Cloning and purification of *P. falciparum* apicoplast TyrRS

The genomic sequence of *Pf*-apitRNA<sup>Tyr</sup> was retrieved from PlasmoDB [20] by sequence homology with the human mitochondrial TyrRS (EAW88518.1, *Hs*-mitoTyrRS) and *Thermus thermophilus* TyrRS (AEG33811.1) [6]. The gene (PF3D7\_1117500) codes for a 561 amino acid protein. The *Pf*-apitRNA<sup>Tyr</sup> gene was amplified by PCR from a *P. falciparum* cDNA library (provided by Dr. H. Vial, Montpellier), sequenced, and cloned into pQE30 (Qiagen) to produce a protein with a 6-histidine fusion tag at its N-terminus. This plasmid expresses the *Pf*-apitRNA<sup>Tyr</sup> recombinant protein without its N-terminal apicoplast targeting signal and covers amino acids 25 to 561 (S1 Fig).

In addition to cloning of the endogenous *P. falciparum* nucleotide sequence of *Pf*-apitRNA<sup>Tyr</sup> (Endo), two additional nucleotide sequences, both encoding the same amino acid chain, were designed (S2 Fig). In the first, we produced an optimized version (Opt) of the *Pf*-apitRNA<sup>Tyr</sup> gene based on *Escherichia coli* codon usage (designed by Genscript) and a harmonized version (Harm) of the gene using the ANACONDA software [21,22]. Both genes were synthesized by GenScript (<https://www.genscript.com>) and cloned into the pQE30 plasmid. Furthermore, a truncated variant of the Harm *Pf*-apitRNA<sup>Tyr</sup> gene, lacking the C-terminal S4-like domain (residues 461 to 561) was cloned into pQE70 with a C-terminal 6-His tag.

Overexpression of all recombinant proteins was performed at 18 °C overnight in LB medium containing 0.1 mg/mL ampicillin and 1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside), and the purification of 6 His-tagged proteins was accomplished on Ni-NTA resin according to the manufacturer's instructions (Qiagen). Purified enzymes were dialyzed against 25 mM HEPES-KOH pH 7.5, 25 mM KCl, 50% glycerol and were kept at -20 °C until use. Proteins were quantified through absorbance measurements and their enzymatic activities were assessed by *in vitro* aminoacylation of native *E. coli* tRNA<sup>Tyr</sup>. Gel filtration analysis was performed on a Superdex 200 increase 10/3000 GL column (GE Healthcare) in 50 mM potassium phosphate buffer pH 7.5, 150 mM KCl, 10% glycerol and 1 mM EDTA.

### Sequence analysis

Sequence alignments were computed with Tcoffee [23], and CLUSTALW [24] softwares. The prediction of *Pf*-apitRNA<sup>Tyr</sup> secondary structure was achieved with the PredictProtein software [25]. The PlasmoAP algorithm [20] confirmed the presence of an apicoplast targeting signal and predicted the cleavage site. The *E. coli* codon usage database was from [26].

### Preparation of tRNA molecules

Native *E. coli* tRNA<sup>Tyr</sup> (Genbank CP010229.1) was purchased from Sigma-Aldrich. Transcripts corresponding to wild-type (PF3D7\_API00700) and mutated *Pf*-apitRNA<sup>Tyr</sup> were synthesized by *in vitro* transcription using the transzyme method [27]. tRNA transcripts were obtained by *in vitro* transcription of synthetic genes cloned in pUC 119 and linearized with *Bst*NI (to produce a 3' CCA sequence after transcription). Each of these genes corresponds to the T7 RNA polymerase promoter, followed by a hammerhead ribozyme, and the tRNA gene sequences. *In vitro* transcriptions were performed in reaction mixtures containing 40 mM

Tris-HCl pH 8.1, 22 mM MgCl<sub>2</sub>, 5 mM dithioerythritol, 0.1 mM spermidine, 4 mM each nucleoside triphosphate, 5 mM GMP, 50 ng/mL linearized plasmid and 5 ng/mL T7 RNA polymerase. Transcription mixtures were incubated for 3 h at 37°C and reactions were stopped by phenol extraction. Auto-catalytically cleaved transcripts correctly ending with the CCA-sequence were purified by preparative electrophoresis on 12% polyacrylamide denaturing gels (8 M urea) followed by electroelution (Schleicher & Schuell apparatus). Concentration of tRNA transcripts was determined by absorbance at 260 nm. We similarly produced transcripts corresponding to the wild-type *P. falciparum* apicoplast tRNA<sup>Ser</sup><sub>GCU</sub> (PF3D7\_API00800, *Pf*-apitRNA<sup>Ser</sup>).

## Procedures for structural analysis of free and TyrRS-complexed tRNAs

Lead and enzymatic probing were performed as in [28] with the following details described here:

**Lead probing:** 1 μM of 5'-labeled *Pf*-apitRNA<sup>Tyr</sup> wild-type transcript (80,000 cpm) was incubated in 50 mM Tris-acetate pH 7.5, 5 mM magnesium acetate, 50 mM potassium acetate. A solution of Pb(OAc)<sub>2</sub>, freshly prepared in H<sub>2</sub>O, was added to reach final concentrations of 1, 3, 6 and 10 mM. The samples were incubated for 6 min at 25 °C.

**Enzymatic probing:** 1 μM of 5'-labeled *Pf*-apitRNA<sup>Tyr</sup> transcript (80,000 cpm) was incubated in 20 mM MgCl<sub>2</sub>, 100 mM NaCl, 50 mM Hepes-NaOH pH 7.5 in the presence of T1 (0.2 U), S1 (5.4 U) and V1 (7x10<sup>-4</sup> U) nucleases for 10 min at 25 °C. For S1 probing, 1 mM ZnCl<sub>2</sub> was added. A control experiment without nuclease was performed in parallel.

All reactions were stopped by the addition of 20 μl of Stop Mix (0.6 M NaOAc, 4 mM EDTA, 0.1 mg/mL total tRNA, and 1 μg glycogen) and ethanol precipitated. The pellets were washed twice with 70% ethanol, vacuum-dried, dissolved in gel loading mix (90% formamide, 0.5% EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue), heated 2 min at 90 °C, and then loaded on a 12% denaturing gel. In parallel, T1 nuclease and alkaline hydrolysis reactions were performed under denaturing conditions to accurately assign the bands in each gel.

Footprinting assays (10 μL) were performed under the same conditions as above in the absence or presence of 5.7 μM *Pf*-apiTyrRS. The tRNA/TyrRS complex was incubated for 6 min at 25 °C, before 0.2 U T1 or 5.4 U S1 (supplemented with 1 mM ZnCl<sub>2</sub> for S1 cleavage) were added. Incubation was continued for 8 min at 25 °C and the reactions were stopped by phenol extraction. After precipitation, the pellets were treated as described above.

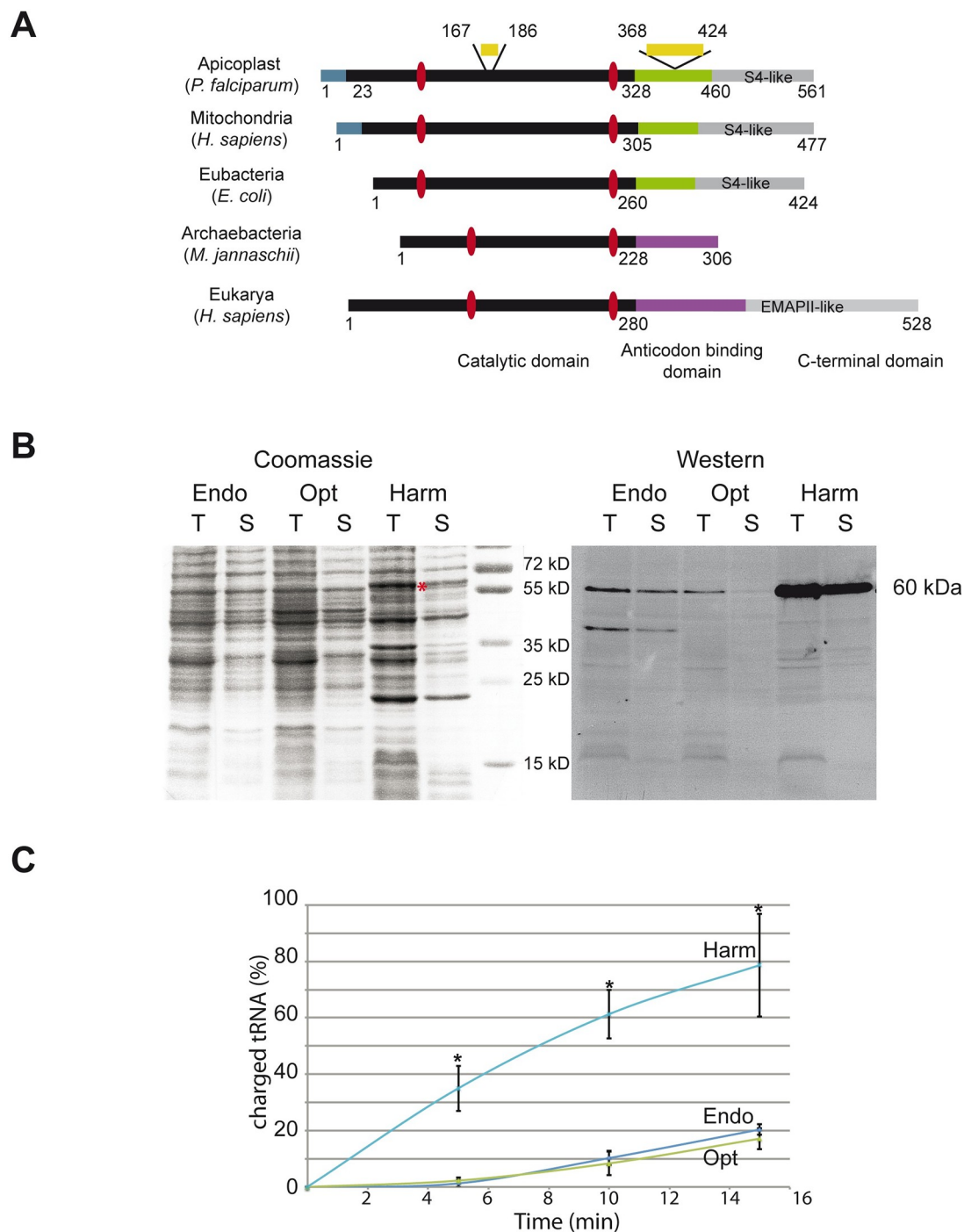
## Aminoacylation assays

Tyrosylation assays were performed at 37 °C in 50 mM Hepes-KOH (pH 7.6), 25 mM KCl, 12 mM MgCl<sub>2</sub>, 2.5 mM ATP, 0.2 mg/ml bovine serum albumin, 1 mM spermine, 10 μM L-[<sup>14</sup>C]-tyrosine at 486 mCi/mmol (Perkin Elmer), and appropriate amounts of either native or transcribed tRNA<sup>Tyr</sup> (0.5 to 10 μM). *Pf*-apiTyrRS (50 to 100 nM, diluted in 100 mM Hepes-NaOH pH 7.4, 1 mM DTT, 5 mg/mL BSA, and 10% glycerol) was added to start the reaction [29]. The apparent kinetic parameters were determined from Lineweaver-Burk plots.

## Results

### Identification of the *P. falciparum* apicoplast TyrRS gene

We identified the gene coding for the 561 amino acid *Plasmodium* *Pf*-apiTyrRS (PF3D7\_1117500) containing an N-terminal catalytic domain, an anticodon binding domain at its C-terminus, and a putative signal sequence for apicoplast targeting [6] (Fig 1A). The Rossman-fold-containing catalytic domain (amino acids 1–328) presents both class I aaRS



**Fig 1. Expression of a soluble and active *Pf*-apiTyrRS in *E. coli*.** (A) The overall organization of *Pf*-apiTyrRS, compared to TyrRSs from humans (mitochondria and cytosol), bacteria, and archaea. Each structural domain is given in a specific color: blue for mitochondrial and apicoplast targeting signals, black for catalytic domains, green for the canonical  $\alpha$ -helical anticodon-binding domains, purple for anticodon-binding domains homologous to TrpRS [31], and grey for additional C-terminal domains (S4-like in bacteria and organelles; EMAPII-like in vertebrates [32]). *Plasmodium*-specific idiosyncratic insertions are shown in yellow. Red dots indicate the position of the signature sequences present in the catalytic domains of class I aaRSs. (B) Coomassie-stained gel (loading control) and the corresponding Western blot show *Pf*-apiTyrRS expressed from the *Plasmodium* wild-type nucleotide sequence (Endo), the optimized gene sequence for *E. coli* expression (Opt), and the harmonized gene sequence (Harm). T stands for Total extract, and S stands for Soluble extract (supernatant of the centrifuged total extract). On the coomassie-stained gel, overexpression of the Harm *Pf*-TyrRS is indicated with a red asterisk. Additional bands potentially correspond to degradation products. On the Western Blot, 6-His tagged proteins were specifically detected with a mouse anti-



penta-his tag antibody (Qiagen). (C) Aminoacylations of *E. coli* tRNA<sup>Tyr</sup> (0.8  $\mu$ M) performed with *Pf*-apiTyrRSs (100 nM) produced from the three gene variants were measured after 5, 10 and 15 min incubation. Controls (without tRNA) were subtracted. Results are an average of three independent experiments. \*significant at  $p < 0.05$  (t-test) (S1 Table).

<https://doi.org/10.1371/journal.pone.0209805.g001>

specific motifs (HNGL and KYSKS). As expected, the C-terminal domain of *Pf*-apiTyrRS presents the typical  $\alpha$ -helical domain (amino acids 329–460) and the S4-like region (amino acids 461–561) that are both specific features found only in bacterial and mitochondrial TyrRSs (reviewed in [30]).

The PlasmoAP algorithm [20] that predicts apicoplast targeting signals indicated that the N-terminal extremity of *Pf*-apiTyrRS "very likely" contains an apicoplast targeting signal and secondary structure predictions positioned this targeting sequence within an  $\alpha$ -helix (S1 Fig). These information led us to the deletion of the first 24 amino acids from the N-terminus in the recombinant *Pf*-apiTyrRS.

*Pf*-apiTyrRS is longer than its prokaryotic homologs because it contains two insertions of 19 and 56 amino acids in the catalytic and the anticodon-binding domains, respectively (Fig 1A). Sequence alignments with seven other *Plasmodium* apicoplast TyrRSs (*P. reichenowi*, *P. vivax*, *P. knowlesi*, *P. gallinaceum*, *P. yoelii*, *P. chabaudi*, *P. berghei*) revealed conserved locations for these insertions, while their sizes and sequences vary significantly (S1 Fig). The insertion located in the anticodon-binding domain is characterized by single amino acid repeats [33]. Indeed, this insertion is composed of 30% asparagine residues.

### Production of a functional recombinant *P. falciparum* apicoplast TyrRS

*Pf*-apiTyrRS<sub>25–561</sub> could be expressed in *E. coli* directly from the *P. falciparum* wild-type nucleotide sequence (Endo), but the affinity-purified yield of soluble protein was poor (<1.5 mg protein per liter culture). Alternative strategies were used to improve the production of soluble *Pf*-apiTyrRS<sub>25–561</sub>. Two different synthetic gene sequences both encoding the same wild-type *Pf*-apiTyrRS<sub>25–561</sub> were cloned (S2 Fig). In one case (Opt), the coding DNA sequence was changed using standard codon optimization rules for expression in *E. coli* (GenScript). In the second case, the DNA sequence was "harmonized" (Harm) using the bioinformatics application for gene primary structure analysis ANACONDA. This program uses statistical methods to analyze not only the codon usage but also the codon context (degree of association, context, and clustering) on a genomic scale [21,22]. In other words, it takes into account the rules governing the evolution of codon bias in *P. falciparum* to design a new nucleotide sequence adapted to the *E. coli* translational apparatus. The main differences between Opt and Harm were at the level of leucine codons, which were all substituted with CTA in the harmonized gene (S2 Fig); CTA is the rarest amongst the six leucine codons (0.385%) in *E. coli* [26].

Gene expression and protein purification using the Opt gene did not change the expression of the protein significantly compared to the endogenous *Plasmodium* DNA sequence (Endo), however, the solubility of *Pf*-apiTyrRS<sub>25–561</sub> was reduced (Fig 1B) and purification yields were low (<0.4 mg protein per liter culture). The best expression yields were obtained with the Harm gene, which yielded nearly 3-fold more *Pf*-apiTyrRS after affinity purification (about 4 mg protein per liter culture) (S3A and S3B Fig). Furthermore, comparative aminoacylation assays (Fig 1C) using native *E. coli* tRNA<sup>Tyr</sup> demonstrated that the enzymes produced from these three plasmids were not functionally equivalent. Indeed, the Harm *Pf*-apiTyrRS is significantly more efficient in aminoacylation than the other two preparations. This observation indicates that gene harmonization not only increased the solubility and hence the purification yields of *Pf*-apiTyrRS<sub>25–561</sub>, but also improved the correct folding of the recombinant protein. Indeed, the purified Harm *Pf*-apiTyrRS<sub>25–561</sub> protein elutes as a major peak of about 160 kDa

on a gel filtration column, suggesting that it forms the expected 118 kDa homodimer (S3C Fig). Thus, *Pf*-apiTyrRS<sub>25-561</sub> expressed from the harmonized construct was used to determine the kinetic parameters of all *Pf*-apitRNA<sup>Tyr</sup> mutants.

*Pf*-apiTyrRS displays an S4-like domain at its C-terminus, specific to bacterial and mitochondrial TyrRSs (Fig 1A). In general, the elimination of this domain increases solubility, while decrease the enzyme's affinity for tRNA<sup>Tyr</sup> (e.g. [34]), since recognition of the tRNA variable region is disrupted [35]. In contrast to these prokaryotic-type TyrRSs, the truncation of the S4-like domain of *Pf*-apiTyrRS led to an inactive enzyme that does not catalyze the first step of the aminoacylation reaction (tyrosine activation in the presence of ATP as measured by ATP/PPi exchange assays (S3D Fig)). This result suggests that, unlike all other known TyrRSs, deletion of the C-terminal domain of *Pf*-apiTyrRS destabilizes the folding of the N-terminal catalytic site.

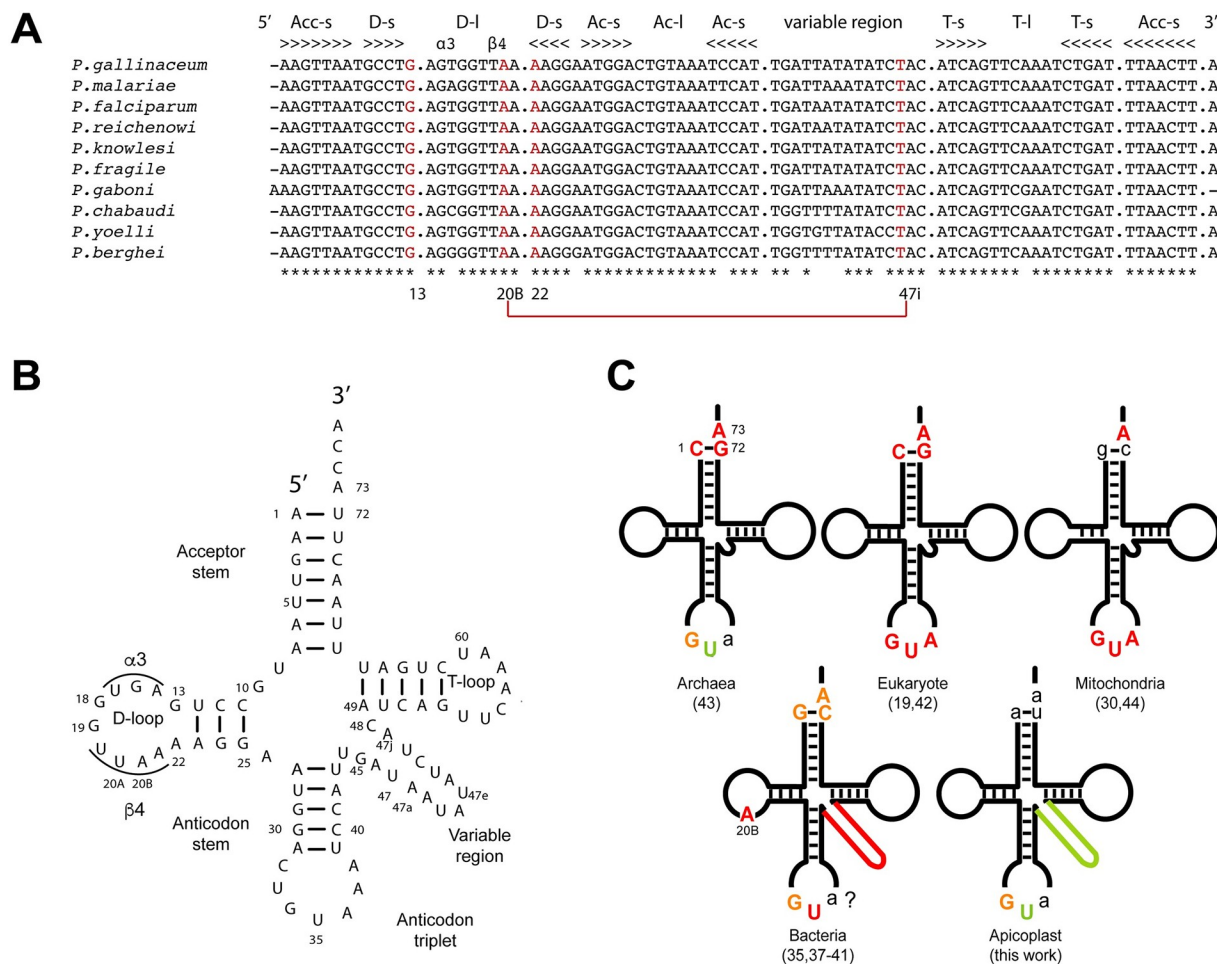
### Sequence peculiarities in *P. falciparum* apicoplast tRNA<sup>Tyr</sup>

The natural substrate for *Pf*-apiTyrRS is the only tRNA<sup>Tyr</sup> encoded by the apicoplast genome (Fig 2A and 2B). *Pf*-apitRNA<sup>Tyr</sup> is composed of 68% A/U residues (both in stems and loops) reflecting the rich A/T composition of the *P. falciparum* genome [3]. *Pf*-apitRNA<sup>Tyr</sup> displays (i) the phylogenetically conserved A<sub>73</sub> discriminator base; (ii) an α3-β4 D-loop with a non-canonical G<sub>13</sub>-A<sub>22</sub> base pair in the D-stem; and (iii) is a class 2 tRNA (like bacterial tRNA<sup>Tyr</sup>s), defined as tRNAs with a large variable region (Fig 2B). Interestingly, *Pf*-tRNA<sup>Tyr</sup> is characterized by the presence of an A<sub>1</sub>-U<sub>72</sub> base pair at the top of the acceptor stem. This base pair is conserved in *Plasmodium* apicoplast tRNA<sup>Tyr</sup> sequences referenced in EupathDB [20] (some of which are displayed in Fig 2A), while tRNA<sup>Tyr</sup> from bacteria, mitochondria, and chloroplasts are characterized by a G<sub>1</sub>-C<sub>72</sub> base pair and tRNA<sup>Tyr</sup> from archaea and eukarya contain a C<sub>1</sub>-G<sub>72</sub> base pair (Fig 2C).

### Cloverleaf folding of *P. falciparum* apicoplast tRNA<sup>Tyr</sup>

The 87 nucleotide *Pf*-apitRNA<sup>Tyr</sup> was produced as a transcript lacking modified bases. Probing experiments were performed with Pb(OAc)<sub>2</sub> and nucleases (Fig 3A) to verify that the high proportion of A and U residues together with the absence of posttranscriptional modifications do not hinder the formation of the canonical cloverleaf fold. S1 and T1 nucleases and Pb(OAc)<sub>2</sub> are specific for single-stranded regions, whereas the V1 nuclease recognizes double-stranded and highly structured regions. In Fig 3A, the similarities between RNase T1 profiles in native (T1) and denaturing (G) conditions suggest that the *Pf*-tRNA<sup>Tyr</sup> transcript is flexible. This hypothesis was confirmed by the pattern of lead cleavage positions, which occur throughout the sequence, with the strongest cuts concentrated in the loops and the variable region.

As expected for a cloverleaf fold, the strongest S1 and V1 cuts are only found in the anticodon-loop and the anticodon-stem, respectively, indicating that this portion of the transcript is indeed folded. Moreover, the moderate accessibilities of A<sub>57</sub> and U<sub>20A</sub> to V1 cleavage confirm the presence of tertiary interactions between the T- and D- loops, whereas conflicting V1 and S1 cuts in the variable region suggest a fluctuating structure in this domain. Altogether, these probing data are in agreement with a cloverleaf fold and indicate intrinsic structural plasticity, reminiscent of what has been observed for some mitochondrial tRNAs [28,45]. Moreover, the correct folding of *Pf*-apitRNA<sup>Tyr</sup> was also confirmed by aminoacylation assays with the homologous Harm *Pf*-apiTyrRS<sub>25-561</sub>, since tyrosylation of *Pf*-apitRNA<sup>Tyr</sup> occurred with the same catalytic efficiency as with the native *E. coli* tRNA<sup>Tyr</sup> (Table 1).



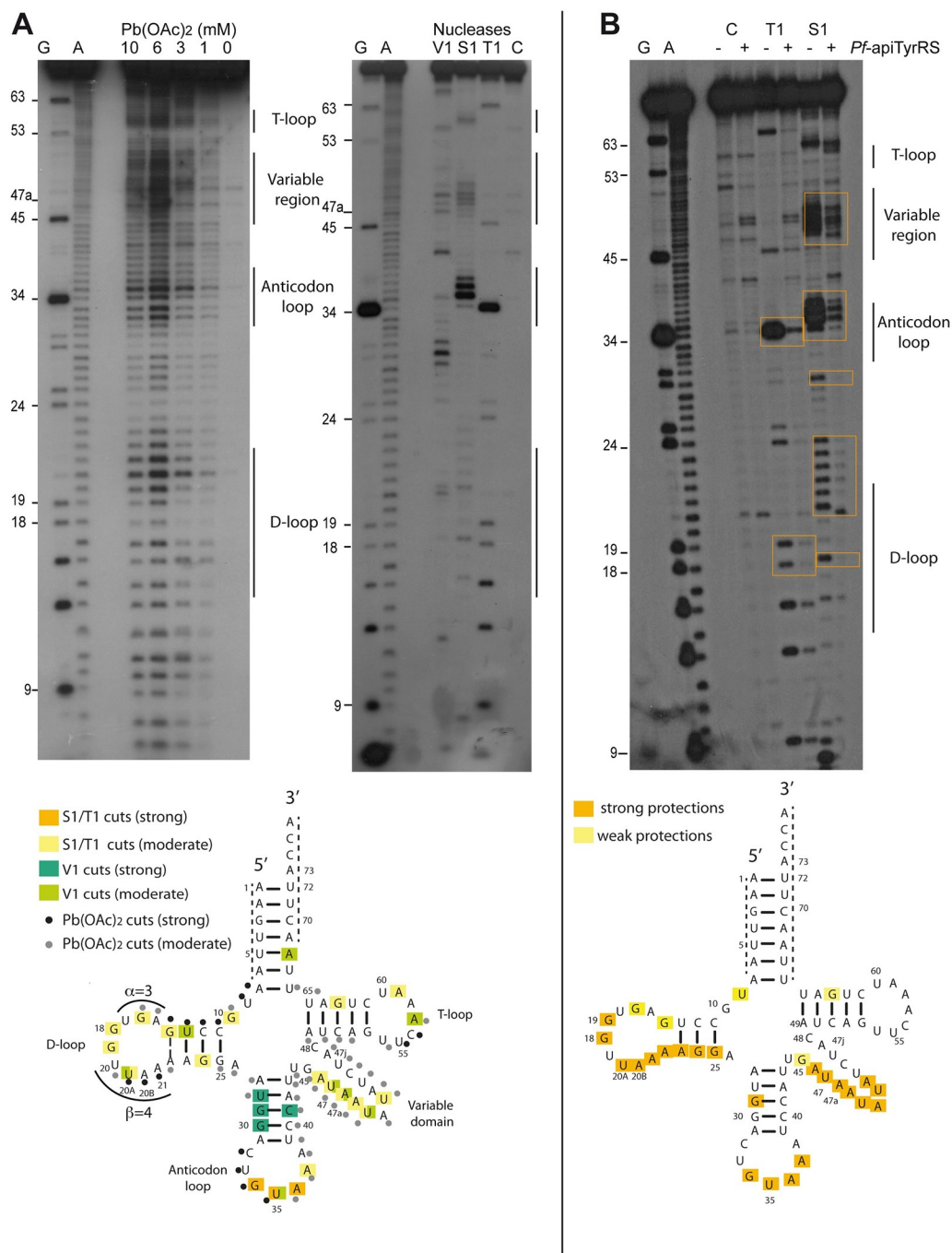
**Fig 2. Sequence of *Pf*-apitRNA<sup>Tyr</sup> and evolution of tyrosine identity.** (A) Comparison of tyrosine-specific apicoplast tRNAs belonging to different *Plasmodium* species. The domains of the cloverleaf structure are indicated at the top: Acc-s refers to the acceptor-stem; D-s refers to the D-stem; D-l refers to the D-loop; Ac-l refers to the anticodon-loop; T-s refers to the T-stem; and T-l refers to the T-loop. Structural elements or nucleotides involved in specific tertiary interactions are highlighted in red. (B) Secondary structure of *Pf*-apitRNA<sup>Tyr</sup>. The architectural organization of the D-loop (α and β domains on both sides of residues G<sub>18</sub>G<sub>19</sub>) is indicated. Notice the non-canonical G<sub>13</sub>-A<sub>22</sub> base pair located at the end of the D-arm. The tRNA is numbered according to [36]. (C) 2D schematic structures of tRNA<sup>Tyr</sup> showing the residues involved in tyrosylation in different phyla [19,30,37–44]. The residues involved in tyrosine identity are explicitly given in uppercase. The strengths of the tyrosine identity elements are indicated by colors: red (loss in catalytic efficiency >100-fold compared to the wild-type transcript), orange (loss between 10- and 100-fold) and green (loss between 5- and 10-fold). Lowercase letters are given to highlight conservation of the residues in the anticodon triplet and at positions 1–72 and 73 in the acceptor stem, despite their exclusion from the identity set. The question mark shows that the importance of position 36 has not been tested in bacteria.

<https://doi.org/10.1371/journal.pone.0209805.g002>

## Extensive recognition of *P. falciparum* apicoplast tRNA<sup>Tyr</sup> by its cognate TyrRS

S1 and V1 nucleases were used in footprinting experiments (Fig 3B) to detect the protected regions of *Pf*-apitRNA<sup>Tyr</sup> in the presence of *Pf*-apiTyrRS<sub>25–561</sub>. The anticodon-loop and the variable domain of the *Pf*-apitRNA<sup>Tyr</sup> transcript are both strongly protected from RNase cleavage and are consistent with what has been observed in the crystallographic structure of the *T. thermophilus* TyrRS/tRNA<sup>Tyr</sup> complex [35]. However, protection patterns detected in the *Pf*-apitRNA<sup>Tyr</sup> D-domain are dissimilar to the bacterial recognition pattern [35]. The addition of *Pf*-apiTyrRS<sub>25–561</sub> protected the *Pf*-apitRNA<sup>Tyr</sup> D-loop from nuclease cleavage, as a





**Fig 3. Probing of *Pf*-apitRNA<sup>Tyr</sup> transcript.** (A) Autoradiograms from probing experiments analyzing 5'-labeled *Pf*-apitRNA<sup>Tyr</sup> transcript with Pb(OAc)<sub>2</sub> (left panel), RNases V1, S1, and T1 (right panel). The final concentrations of Pb(OAc)<sub>2</sub> are indicated at the top of the autoradiogram. The G lane corresponds to a denaturing RNase T1 ladder and lane A to an alkaline ladder. The C lane shows the control experiment performed with untreated transcript. The locations of D-, anticodon- and T- loops, as well as the large variable domain, are noted. Results are summarized on the *Pf*-apitRNA<sup>Tyr</sup> cloverleaf structure. The tRNA is numbered according to [36]. Nucleotides at the 3'- and 5'- ends, which cannot be analyzed are indicated by dotted lines on the tRNA structure. (B) Autoradiogram corresponding to footprinting experiments: the 5'-labeled *Pf*-apitRNA<sup>Tyr</sup> transcript was incubated with S1 and T1 nucleases in the absence (-) or presence (+) of *Pf*-apiTyrRS. The strongest RNase protections which confirm the sites of interaction between the tRNA<sup>Tyr</sup> transcript and *Pf*-apiTyrRS are framed in orange. Controls (C) were performed without nucleases. G and A indicate T1 and alkaline ladders, respectively. *Pf*-apitRNA<sup>Tyr</sup> residues that are protected from nucleases in the presence of *Pf*-apiTyrRS are indicated in orange (strong protection) and yellow (weak protection) on the *Pf*-apitRNA<sup>Tyr</sup> cloverleaf structure.

<https://doi.org/10.1371/journal.pone.0209805.g003>

**Table 1. Tyrosylation kinetic parameters of *E. coli* tRNA<sup>Tyr</sup> and wild-type or mutated *Pf*-apitRNAs by *Pf*-apiTyrRS.**

	$K_m$ ( $\mu M$ )	S.D. $K_m$	$k_{cat}$ ( $10^{-3} sec^{-1}$ )	S.D. $k_{cat}$	$k_{cat}/K_m$ ( $10^{-3} sec^{-1} \cdot \mu M^{-1}$ )	p-value	L <sup>a</sup> (x-fold)
<i>E. coli</i> tRNA <sup>Tyr</sup>	0.9	0.2	12.0	5.1	14.1	0.197	1.5
<i>Pf</i> -apitRNA <sup>Tyr</sup>	0.8	0.3	16.0	4.0	20.6		1.0
<i>Pf</i> -apitRNA <sup>Ser</sup>	nd		nd		-	-	>1000
<b>Acceptor stem</b>							
G <sub>73</sub>	1.0	0.3	9.8	5.4	9.2	0.021	2.2*
C <sub>73</sub>	0.5	0.2	4.1	0.5	8.9	0.147	2.3
U <sub>73</sub>	0.8	0.3	4.0	1.7	4.9	0.019	4.2
G <sub>1</sub> -C <sub>72</sub>	1.2	1.0	10.3	3.0	17.4	0.793	1.2
C <sub>1</sub> -G <sub>72</sub>	0.3	0.7	4.5	2.3	10.8	0.007	1.9*
C <sub>2</sub> -G <sub>71</sub> /U <sub>3</sub> -A <sub>70</sub>	0.4	0.2	3.0	0.6	8.0	0.019	2.6*
<b>Anticodon triplet</b>							
C <sub>34</sub>	6.2	3.4	3.2	1.6	0.6	0.014	36.0*
G <sub>35</sub>	0.7	0.2	13.3	6.5	19.8	0.786	1.0
C <sub>35</sub>	1.6	0.5	4.2	1.6	2.7	0.012	7.7*
A <sub>35</sub>	1.7	0.7	4.5	0.8	3.0	0.008	6.8*
C <sub>36</sub>	0.6	0.3	3.5	0.8	6.8	0.015	3.0*
U <sub>36</sub>	1.0	0.3	6.7	1.5	7.0	0.021	3.0*
G <sub>36</sub>	1.6	0.4	9.7	2.2	7.0	0.026	3.0*
Ser (C <sub>35</sub> U <sub>36</sub> )	1.3	0.3	1.3	0.4	1.1	0.014	19.1*
<b>Variable region (Vr)</b>							
ΔVr	1.8	1.0	4.5	0.9	2.9	0.014	7.1*
SerVr	0.8	0.4	5.6	2.3	7.6	0.013	2.7*
<b>D-loop (D-l)</b>							
U <sub>20B</sub>	0.9	0.4	8.2	3.0	12.6	0.353	1.6
SerVr+D-l	1.0	0.3	7.4	2.4	8.1	0.047	2.1*

$K_m$  and  $k_{cat}$  values correspond to means from three independent experiments or more (Table S1) with the indicated error (S.D.). Mean values of  $k_{cat}/K_m$  and p-values were calculated independently.

\*significant at  $p < 0.05$  (t-test).

<sup>a</sup>L values correspond to losses of catalytic efficiency relative to wild-type *Pf*-apitRNA<sup>Tyr</sup>. nd indicates not detectable.

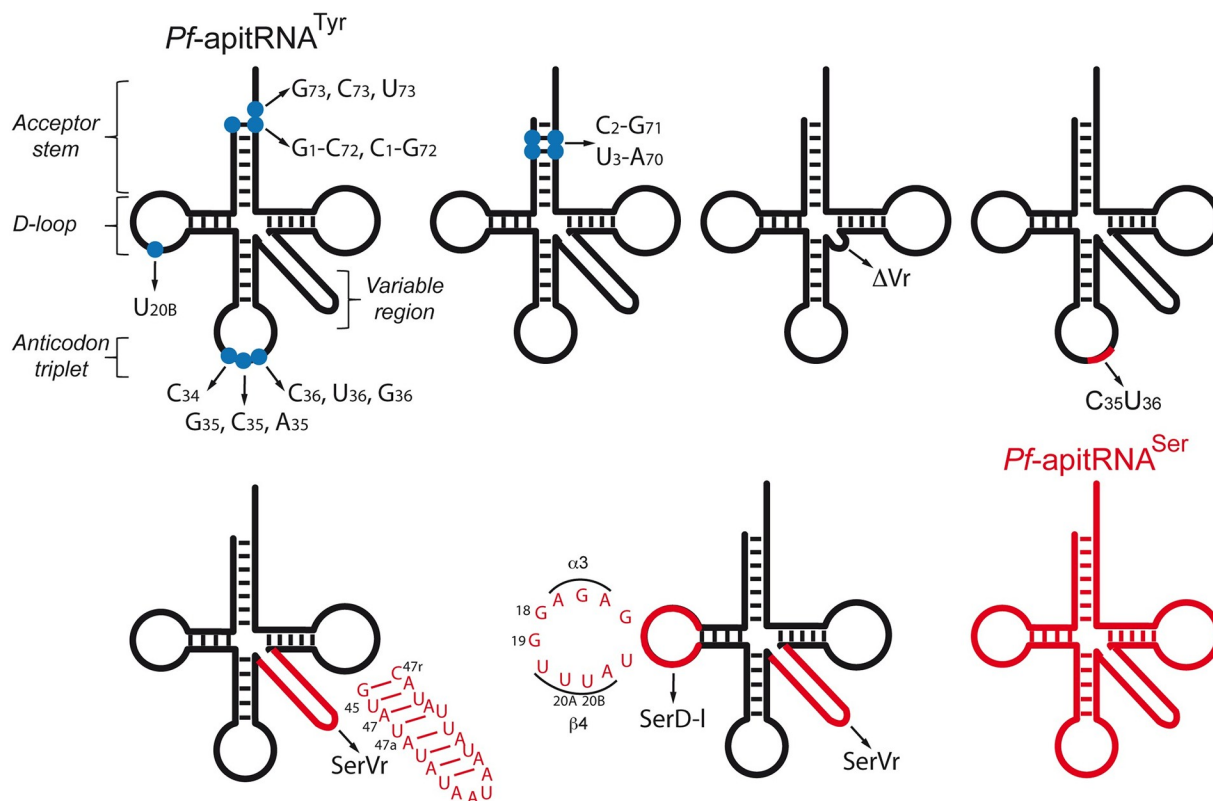
<https://doi.org/10.1371/journal.pone.0209805.t001>

consequence either of a direct contact with the synthetase, or of an indirect effect due to a conformational change in the tRNA bound to the enzyme.

### Looking for identity elements in *P. falciparum* apicoplast tRNA<sup>Tyr</sup>

TyrRS displays species-specific tRNA recognition (summarized in Fig 2C with the references therein). The tRNA<sup>Tyr</sup> A<sub>73</sub> discriminator base and the G<sub>34</sub> anticodon nucleotide are universally important for TyrRS recognition. Anticodon nucleotides U<sub>35</sub> and A<sub>36</sub> also contribute to tyrosylation identity, yet with varying strengths in eukarya, bacteria, and mitochondria, and are marginal in archaea. Notably, the G<sub>1</sub>-C<sub>72</sub> identity base-pair, located at the top of the acceptor stem in bacterial tRNA<sup>Tyr</sup>, is replaced by a C<sub>1</sub>-G<sub>72</sub> identity base pair in eukarya tRNA<sup>Tyr</sup>. Finally, the large variable region is unique to bacterial tRNA<sup>Tyr</sup> and is essential for tyrosylation.

We chose to elucidate the identity determinant set for *Pf*-apitRNA<sup>Tyr</sup> with *Pf*-api-TyrRS<sub>25-561</sub>. Eighteen mutants were designed to test the extremity of the acceptor stem (discriminator base 73 and the 1–72 base-pair), the anticodon triplet, and the variable region for their importance



**Fig 4.** 2D schematic structures of *Pf*-apitRNA<sup>Tyr</sup> summarizing the locations and the nature of the mutated residues. Mutated sequences are indicated by blue dots and mutations are listed. In addition to the single mutations shown, 10 residues were deleted from the variable region ( $\Delta V_r$ : from 47 to 47i). The 2D schematic structure of *Pf*-apitRNA<sup>Ser</sup> and the serine-specific features introduced into *Pf*-apitRNA<sup>Tyr</sup>, such as dinucleotide C<sub>35</sub>U<sub>36</sub>, the variable region (SerVr), and the D-loop plus the variable region (SerVr+D-I) are shown in red.

<https://doi.org/10.1371/journal.pone.0209805.g004>

for tyrosylation (Fig 4). The kinetic parameters for tyrosylation were determined and compared to those obtained for the wild-type *Pf*-apitRNA<sup>Tyr</sup> transcript (Table 1).

**Acceptor stem.** Unexpectedly, replacement of the A<sub>73</sub> discriminator base by G, C or U reduced tyrosylation only 2.2 to 4.2-fold, suggesting that these mutations do not affect the recognition of tRNA<sup>Tyr</sup> by *Pf*-apiTyrRS, but slightly modulate the structure of the tRNA near the catalytic site. Indeed, it has been shown that a pyrimidine residue at position 73 might influence the stability of the acceptor stem extremity [46]. Unlike the vast majority of tRNA<sup>Tyr</sup> isoacceptors, where the first base pair in the acceptor stem is G<sub>1</sub>-C<sub>72</sub> (in bacteria) or C<sub>1</sub>-G<sub>72</sub> (in archaea and eukarya), all *Plasmodium* apicoplast tRNA<sup>Tyr</sup> contain a conserved A<sub>1</sub>-U<sub>72</sub> (Fig 2A). A<sub>1</sub>-U<sub>72</sub> was therefore changed to G-C and C-G, (Fig 4 and Table 1). *Pf*-apiTyrRS<sub>25-561</sub> aminoacylates the wild-type transcript, the G<sub>1</sub>-C<sub>72</sub> and the C<sub>1</sub>-G<sub>72</sub> variants with similar kinetic values, indicating that the first base pair of the acceptor stem is not part of this system's tyrosine identity set, providing the second example (with the human mitochondrial TyrRS) of a TyrRS lacking specificity to base pair 1–72 [44]. We hypothesized that a different base pair in the acceptor stem might have replaced 1–72 in the identity set, so we mutated positions 2–71 and 3–70 (Fig 4 and Table 1). The recombinant *Pf*-apiTyrRS aminoacylated these variants with kinetic parameters similar to those of the wild-type transcript, characterized by a loss in efficiency of only 2.6-fold. Together, these data show the absence of tyrosine identity elements in the acceptor stem and suggest that *Pf*-apiTyrRS only recognizes the ribose-phosphate backbone in this region of the tRNA molecule.

**Anticodon triplet.** Residues 34, 35 and 36 were individually mutated in *Pf*-apitRNA<sup>Tyr</sup>. Only the variant where C<sub>34</sub> replaced G<sub>34</sub> showed a substantial loss in tyrosylation (36-fold). Mutations at position 35 affected catalytic efficiency by 1 to 7.7-fold and mutations at position 36 had virtually no effect on activity (3.0-fold) (Table 1). Moreover, converting the tyrosine anticodon to a serine GCU anticodon (mutant Ser C<sub>35</sub>U<sub>36</sub>) only leads to a moderate but significant decrease of 19.1-fold (Table 1).

**Variable region and D-loop.** Shortening the variable region ( $\Delta$ Vr) decreased the catalytic efficiency by a factor 7.1, compared to wild-type *Pf*-apitRNA<sup>Tyr</sup> (Table 1). However, insertion of the variable region of *Pf*-apitRNA<sup>Ser</sup> (SerVr) reduced this effect to only a 2.7-fold reduction in activity, suggesting that our  $\Delta$ Vr deletion caused a change in tRNA structure rather than a direct impact on enzyme recognition. In *T. thermophilus* tRNA<sup>Tyr</sup>, A<sub>20B</sub>, located in the D-loop, interacts with U<sub>47i</sub> in the variable domain, which provides a precise orientation of the variable region for its optimal recognition by TyrRS [35]. Since the *Pf*-apitRNA<sup>Tyr</sup> sequence displays both A<sub>20B</sub> and U<sub>47i</sub>, the same tertiary interaction could form (Fig 3A). Replacement of A<sub>20B</sub> by U<sub>20B</sub> in the *Pf*-apitRNA<sup>Tyr</sup> D-loop should, therefore, eliminate this interaction. However, this mutant showed no loss in tyrosylation activity (1.6-fold) (Table 1). Finally, both the D-loop and the *Pf*-apitRNA<sup>Tyr</sup> variable region were changed to the *Pf*-apitRNA<sup>Ser</sup> (SerVr+D-l) sequences. In contrast to the long variable region in *Pf*-apitRNA<sup>Tyr</sup>, the long variable region of *Pf*-apitRNA<sup>Ser</sup><sub>GCU</sub> exhibits eight base pairs and potentially a specific tertiary interaction between G<sub>45</sub>-C<sub>48a</sub> and U<sub>20B</sub>, determining its spatial orientation as in *T. thermophilus* [35] (S4 Fig). This replacement had no significant effect on tyrosylation efficiency (2.1-fold). Together these mutants demonstrate that, in the apicoplast, the presence of a long variable region plays a weak but significant role in tyrosylation identity ( $\Delta$ Vr, 7.1-fold), but neither the sequence nor the orientation of this long variable region is involved.

## Discussion

### Expression of *Pf*-apiTyrRS

*Plasmodium* aaRSs are longer than their homologs because they contain many peculiar, sequence-repetitive insertions [3]. Neither the synthesis nor the functions of these insertions are understood [33]; the presence of long single amino acid repeats often reduces the solubility of the recombinant protein, but they cannot be removed under penalty of obtaining an inactive protein (e.g., [47]). Insertions are more frequent and more extended in apicoplast than in cytosolic *Plasmodium* aaRSs. Moreover, it has been suggested that the translation of these additional sequences locally reduces the rate of ribosomes and could be used to regulate co-translational folding of proteins [6]. The presence of these insertions and the challenges they introduce may explain why apicoplast aaRSs are poorly studied despite their interest for the development of new anti-malarial drugs [13–15]. To date, only four apicoplast aaRSs from *P. falciparum* have been cloned, expressed, and characterized, namely lysyl- (LysRS) [48], glutamyl- (GluRS) [49], tryptophanyl- (TrpRS) [15], and the dual-targeted cysteinyl-tRNA (CysRS) synthetases [8]. CysRS and GluRS do not contain insertions, thus LysRS and TrpRS are the only insertion-containing apicoplast aaRSs that have been studied to date.

Translation is influenced by the choice of synonymous codons, which specify the same amino acid but differ in their decoding properties [50]. Thus, the primary structure of mRNA contains information that affects translation efficiency. The dominant model is that some codons or codon combinations reduce the decoding rate of ribosomes and thereby isolate the synthesis and folding of well-defined protein domains (e.g. [51–53]). The availability of tRNAs that decode synonymous codons, their requirement for wobble decoding, as well as interactions between adjacent codons play fundamental roles in this model. Codon usage and the

number of tRNA genes in *Plasmodia* are very different from those of *E. coli* and thus make the expression of *Plasmodium* multidomain proteins challenging in the *E. coli* heterologous expression system. In this study, expression of recombinant *Pf*-apiTyrRS directly from the *Plasmodium* mRNA sequence was indeed ineffective. The optimization of synonymous codons involves the selection of optimal codons decoded only by abundant tRNAs in the expression host and thus the simultaneous minimization of rare codons [50,54]. This approach further reduced the solubility of the produced *Pf*-apiTyrRS. However, the use of harmonized codons, designed by the ANACONDA algorithm [21,22], increased the synthesis, solubility, and enzymatic activity of the purified recombinant protein. The main difference between both the Endo and the Opt genes compared to the Harm gene is the systematic replacement of leucine codons (TTA, TTG, CTG, CTT, and CTC) with the rarest leucine codon used in *E. coli* translation (CTA). As the ribosome slows when it encounters rare codons it may help the protein to fold appropriately, thereby increasing the yield of soluble proteins. Here, the Harm gene, containing a combination of fast and slow codons, facilitates co-translational folding and thus the production of a biologically active *Pf*-apiTyrRS. This result suggests that such an approach could be used to overcome the difficulties encountered when expressing *Plasmodium* multidomain proteins.

## Evolution of tyrosine identity

Experimental work on tyrosylation systems from different species has established the evolution of the tyrosine identity set (Fig 2C). The A<sub>73</sub> discriminator base and the G<sub>34</sub> and U<sub>35</sub> anticodon bases were determined as common identity elements in tRNA<sup>Tyr</sup> of bacteria, archaea, eukarya, and mitochondria (summarized in [30]). In addition, the 1–72 base-pair at the end of the acceptor stem is critical to archaeal and eukaryal tyrosylation systems, whereas the long variable region is required only for the correct recognition of bacterial tRNAs (Fig 2C).

From our mutational analysis, the *Plasmodium* apicoplast tyrosylation system retains only one moderate (G<sub>34</sub>) and two weak (U<sub>35</sub> and the long variable region) identity elements to ensure specific aminoacylation. Unlike other tyrosylation systems, this identity set does not include residues in the tRNA acceptor arm. Indeed, despite its conservation in all *Plasmodium* apitRNA<sup>Tyr</sup>, the first A<sub>1</sub>-U<sub>72</sub> base pair was not involved in tyrosylation, a situation already observed for the human mitochondrial tyrosine system [44]; and the A<sub>73</sub> discriminator base, common to all tyrosylation identity sets, does not influence tyrosylation in the *Plasmodium* apicoplast. Besides, neither the structure nor the orientation of the variable region is sufficient to prevent apicoplast tyrosylation. In the *Plasmodium* apicoplast system, insertion of the tRNA<sup>Ser</sup> variable region into *Pf*-apitRNA<sup>Tyr</sup> (mutants SerVr and SerVr+D-I) does not significantly reduce its recognition by *Pf*-apiTyrRS (2.7 and 2.1-fold, Table 1), while swapping the sequence of the variable region of *E. coli* tRNA<sup>Tyr</sup> with that of *E. coli* tRNA<sup>Ser</sup> decreases tyrosylation by more than 300-fold [40].

The only critical effect in the anticodon was obtained when G<sub>34</sub> was mutated, which led to a loss in efficiency of only 36-fold; an unprecedented situation amongst tyrosylation systems. However, the identity of tRNAs is not only dictated by the presence of sets of positive identity elements allowing recognition by cognate synthetases, but also by negative signals that prevent the interaction of tRNAs with non-cognate synthetases. This scenario could play an important role in the *Plasmodium* apicoplast. Of the 27 tRNA gene sequences encoded by this genome, four contain a G<sub>34</sub>T<sub>35</sub> sequence (*Pf*-apitRNA<sup>Tyr</sup><sub>GTA</sub>, tRNA<sup>Asn</sup><sub>GTT</sub>, tRNA<sup>Asp</sup><sub>GTC</sub>, tRNA<sup>His</sup><sub>GTG</sub>) and two contains a G<sub>34</sub> and a long variable region (*Pf*-apitRNA<sup>Tyr</sup><sub>GTA</sub> and tRNA<sup>Ser</sup><sub>GCT</sub>) (S4 Fig). Thus, the non-cognate asparagine, aspartate, histidine, and serine tRNAs must display features prohibiting recognition and tyrosylation by *Pf*-apiTyrRS. The transcription method



used in the present study yields tRNAs lacking modified nucleotides, which may be a disadvantage if post-transcriptional modifications of native tRNAs play such a negative role in identity [17]. We can only predict that some modifications may be present in the apicoplast when the modification enzymes have been annotated in the *Plasmodium* genome [20]. For example, three putative queueine tRNA-ribosyltransferase are found in EupathDB, one of which (PF3D7\_1242200) is predicted to be targeted to the apicoplast. Queueosine and its derivatives are found in bacterial and eukaryal tRNAs with a G34 [55], and guarantee fidelity and efficiency of translation [56]. The presence of this modifying enzyme in the apicoplast suggests that local tyrosine, histidine, aspartate, and asparagine tRNAs can be modified at position 34. However, nothing is known about post-transcriptional modifications of *Plasmodium* apicoplast tRNAs and if idiosyncratic modification patterns can control aminoacylation specificities.

We propose that the high A/T content of the *Plasmodium* apicoplast genome significantly reduces the potential for identity nucleotide combinations in apicoplast tRNAs. In the specific case of *Pf*-apitRNA<sup>Tyr</sup>, this led to the conservation of a minimal identity set with only three weak identity features positively recognized by the *Pf*-apiTyrRS. It is reasonable to ask whether these elements are sufficient to drive tyrosylation *in vivo* efficiently. Tyrosylation specificity could be mainly maintained by the presence of negative determinants (sequence/structural features or post-transcriptional modifications), which prevent mischarging of other *Pf*-apitRNAs by *Pf*-apiTyrRS.

## Supporting information

**S1 Fig. Multiple sequence alignments of eight *Plasmodium* apiTyrRSs and comparison with the human mitochondrial TyrRS.** Protein sequences are from EupathDB: *P. falciparum*\_3D7 (PF3D7\_1117500), *P. reichenowi*\_CDC (PRCDC\_1115900), *P. vivax*\_P01 (PVP01\_0918100), *P. knowlesi*\_strain\_H (PKNH\_0915200), *P. gallinaceum*\_8A (PGAL8A\_00344100), *P. yoelii*\_yoelii\_YM (PYYM\_0931900), *P. chabaudi*\_chabaudi (PCHAS\_0913800) and *P. berghei*\_ANKA (PBANKA\_0930500). The color code follows that of Fig 1A: residues belonging to the catalytic domain are in black with class I signature motifs highlighted in red; residues from the anticodon-binding domain are in green; the S4-like domain is in grey; and the two *Plasmodium*-specific insertions are in yellow. The starting position of recombinant *Pf*-apiTyrRS<sub>25-561</sub> is indicated in cyan. Alignments were performed with Tcoffee [23] and predicted  $\beta$ -sheets and  $\alpha$ -helices of *Pf*-apiTyrRS predicted by the PredictProtein software [20] are indicated with green arrows and red rectangles, respectively.  
(DOCX)

**S2 Fig. DNA sequences encoding *Pf*-apiTyrRS.** Alignment of nucleotide sequences corresponding to the endogeneous (Endo), optimized (Opt) and harmonized (Harm) gene sequences encoding *Pf*-apiTyrRS. The amino acid sequence of the protein is in bold. All leucine (L) codons are highlighted: red indicates codons whose usage in *E. coli* is higher than 1% (TTG, TTA, CTG, CTT and CTC) and in green for the only rare leucine codon (CTA, 0.385%) [26].  
(DOCX)

**S3 Fig. Purification of *Pf*-apiTyrRSs and ATP/PPi exchange assays.** (A) SDS-PAGE of purified *Pf*-apiTyrRS Endo, Opt and Harm (55 and 72 kDa protein molecular weight markers are indicated). (B) Comparison of the purification yields of *Pf*-apiTyrRS expressed from Endo, Opt and Harm genes (relative to Endo); two independent purifications are shown (dark and light grey). (C) Gel filtration profile of *Pf*-apiTyrRS<sub>25-561</sub> Harm (blue); gel filtration standards

(grey) correspond to (1) thyroglobin (670 kDa), (2) bovine  $\gamma$ -globine (158 kDa), (3) chicken ovalbumine (44 kDa), (4) equine myoglobine (17 kDa) and (5) vitamine B-12 (1.35 kDa). (D) ATP/PPi exchange assays. The experiments were performed according to [57], in the presence of tyrosine (2 mM) and radiolabeled [<sup>32</sup>P]Ppi (20 cpm/pmol, this high specific activity was used to detect low exchange activities) and 0.8  $\mu$ M *Pf*-apiTyrRS<sub>25-561</sub> (squares) or the *Pf*-apiTyrRS<sub>25-460</sub> deprived of its S4-like domain (spheres). A negative control with *Pf*-apiTyrRS<sub>25-561</sub> without tyrosine was done in parallel (triangles). (TIF)

**S4 Fig. Sequences of *P. falciparum* apicoplast tRNAs.** Structural alignment of *P. falciparum* apitRNA genes [58] with accession numbers. The anticodon sequences are indicated in red. The domains of the cloverleaf structure are indicated at the top: Acc-s: acceptor-stem; D-s: D-stem; D-l: D-loop; Ac-s: anticodon-stem; Ac-l: anticodon-loop; T-s: T-stem; and T-l: T-loop. Note the presence of introns in the tRNA<sup>Leu</sup><sub>TAA</sub> and potentially in tRNA<sup>Met</sup><sub>CAT-2</sub> gene sequences. On the basis of their sequences, we could not assign initiator versus elongator functions to the two tRNA<sup>Met</sup> isoacceptors. (DOCX)

**S1 Table. Raw data values obtained in aminoacylation experiments.** (A) Data used in Fig 1C correspond to the percentage of aminoacylated tRNAs after 5, 10 and 15 min incubation, in three independent experiments in the presence of Endo, Opt and Harm *Pf*-apiTyrRS. Corresponding means and errors (S.D.) are given and the p-values (t-test) were calculated for the Harm *Pf*-apiTyrRS compared to Endo (p-value(Endo)) or Opt (p-value(Opt)) *Pf*-apiTyrRSs. Significant p-values at  $p < 0.05$  are indicated in red. (B, C) Data corresponds to the kinetic parameters presented in Table 1. The individual  $K_m$  and  $k_{cat}$  values used to calculate means, errors (S.D.) and p-values are indicated. These apparent kinetic parameters were determined from Lineweaver-Burk plots. Significant p-values at  $p < 0.05$  (t-test) are shown in red and Loss values correspond to losses of catalytic efficiency relative to wild-type *Pf*-apitRNA<sup>Tyr</sup>. (PDF)

## Acknowledgments

We are grateful to Tamara Hendrickson and Richard Giegé for critical reading of the manuscript and Anne Théobald-Dietrich for technical assistance. This work was supported by the CNRS, the Université de Strasbourg and the Fondation pour la Recherche Médicale (FRM) [grant numbers FDT201704337050]. This work has been published under the framework of the LABEX: ANR-10-LABX-0036\_NETRINA.

## Author Contributions

**Conceptualization:** Magali Frugier, Joëlle Rudinger-Thirion.

**Data curation:** Joëlle Rudinger-Thirion.

**Funding acquisition:** Magali Frugier.

**Investigation:** Marta Cela, Caroline Paulus, Joëlle Rudinger-Thirion.

**Methodology:** Caroline Paulus.

**Resources:** Magali Frugier.

**Software:** Gabriela R. Moura.

**Supervision:** Joëlle Rudinger-Thirion.

**Validation:** Joëlle Rudinger-Thirion.

**Writing – original draft:** Joëlle Rudinger-Thirion.

**Writing – review & editing:** Manuel A. S. Santos, Magali Frugier.

## References

1. Jackson KE, Habib S, Frugier M, Hoen R, Khan S, Pham JS, et al. Protein translation in *Plasmodium* parasites. Trends Parasitol. 2011; 27(10):467–76. <https://doi.org/10.1016/j.pt.2011.05.005> PMID: 21741312
2. Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, et al. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol. 2004; 2(3):203–16. <https://doi.org/10.1038/nrmicro843> PMID: 15083156
3. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002; 419:498–511. <https://doi.org/10.1038/nature01097> PMID: 12368864
4. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, et al. Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. Science. 2003; 299(5607):705–8. <https://doi.org/10.1126/science.1078599> PMID: 12560551
5. Ibba M, Söll D. Aminoacyl-tRNA synthesis. Annu Rev Biochem. 2000; 69:617–50. <https://doi.org/10.1146/annurev.biochem.69.1.617> PMID: 10966471
6. Frugier M, Bour T, Ayach M, Santos MA, Rudinger-Thirion J, Théobald-Dietrich A, et al. Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. FEBS Lett. 2010; 584(2):448–54. <https://doi.org/10.1016/j.febslet.2009.11.004> PMID: 19900443
7. Jackson KE, Pham JS, Kwek M, De Silva NS, Allen SM, Goodman CD, et al. Dual targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in *Plasmodium falciparum*. Int J Parasitol. 2012; 42(2):177–86. <https://doi.org/10.1016/j.ijpara.2011.11.008> PMID: 22222968
8. Pham JS, Sakaguchi R, Yeoh LM, De Silva NS, McFadden GI, Hou YM, et al. A dual-targeted aminoacyl-tRNA synthetase in *Plasmodium falciparum* charges cytosolic and apicoplast tRNA<sup>Cys</sup>. Biochem J. 2014; 458(3):513–23. <https://doi.org/10.1042/BJ20131451> PMID: 24428730
9. Sharma A, Sharma A. *Plasmodium falciparum* mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. Biochem J. 2015; 465(3):459–69. <https://doi.org/10.1042/BJ20140998> PMID: 25391660
10. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature. 1990; 347(6289):203–6. <https://doi.org/10.1038/347203a0> PMID: 2203971
11. Cusack S, Berthet-Colominas C, Hartlein M, Nassar N, Leberman R. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. Nature. 1990; 347(6290):249–55. <https://doi.org/10.1038/347249a0> PMID: 2205803
12. Huot JL, Enkler L, Megel C, Karim L, Laporte D, Becker HD, et al. Idiosyncrasies in decoding mitochondrial genomes. Biochimie. 2014; 100:95–106. <https://doi.org/10.1016/j.biochi.2014.01.004> PMID: 24440477
13. Khan S. Recent advances in the biology and drug targeting of malaria parasite aminoacyl-tRNA synthetases. Malar J. 2016; 15:203. <https://doi.org/10.1186/s12936-016-1247-0> PMID: 27068331
14. Saint-Léger A, Sinadinou C, Ribas de Pouplana L. The growing pipeline of natural aminoacyl-tRNA synthetase inhibitors for malaria treatment. Bioengineered. 2016; 7(2):60–4. <https://doi.org/10.1080/21655979.2016.1149270> PMID: 26963157
15. Pasaje CF, Cheung V, Kennedy K, Lim EE, Baell JB, Griffin MD, et al. Selective inhibition of apicoplast tryptophanyl-tRNA synthetase causes delayed death in *Plasmodium falciparum*. Sci Rep. 2016; 6:27531. <https://doi.org/10.1038/srep27531> PMID: 27277538
16. Giegé R, Springer M. Aminoacyl-tRNA Synthetases in the Bacterial World. EcoSal Plus. 2016; 7(1).
17. Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. 1998; 26(22):5017–35. PMID: 9801296
18. Giegé R, Eriani G. Transfer RNA recognition and aminoacylation by synthetases. eLS. 2014.

19. Wakasugi K, Quinn CL, Tao N, Schimmel P. Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant. *EMBO J*. 1998; 17(1):297–305. <https://doi.org/10.1093/emboj/17.1.297> PMID: 9427763
20. Aurrecochea C, Barreto A, Basenko EY, Brestelli J, Brunk BP, Cade S, et al. EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Res*. 2017; 45(D1):D581–91. <https://doi.org/10.1093/nar/gkw1105> PMID: 27903906
21. Pinheiro M, Afreixo V, Moura G, Freitas A, Santos MA, Oliveira JL. Statistical, computational and visualization methodologies to unveil gene primary structure features. *Methods Inf Med*. 2006; 45(2):163–8. PMID: 16538282
22. Moura GR, Pinheiro M, Freitas A, Oliveira JL, Frommlet JC, Carreto L, et al. Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. *PLoS One*. 2011; 6(10):e26817. <https://doi.org/10.1371/journal.pone.0026817> PMID: 22046369
23. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, et al. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res*. 2006; 34(Web Server issue):W604–8. <https://doi.org/10.1093/nar/gkl092> PMID: 16845081
24. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010; 38(Web Server issue):W695–9. <https://doi.org/10.1093/nar/gkq313> PMID: 20439314
25. Rost B, Yachdav G, Liu J. The PredictProtein server. *Nucleic Acids Res*. 2004; 32(Web Server issue):W321–6. <https://doi.org/10.1093/nar/gkh377> PMID: 15215403
26. OpenWetWare contributors. Escherichia coli/Codon usage [Internet]. OpenWetWare. 2014.
27. Fechter P, Rudinger J, Giegé R, Théobald-Dietrich A. Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. *FEBS Lett*. 1998; 436(1):99–103. PMID: 9771901
28. Bonnefond L, Florentz C, Giegé R, Rudinger-Thirion J. Decreased aminoacylation in pathology-related mutants of mitochondrial tRNA<sup>Tyr</sup> is associated with structural perturbations in tRNA architecture. *RNA*. 2008; 14(4):641–8. <https://doi.org/10.1261/rna.938108> PMID: 18268021
29. Sohm B, Frugier M, Brulé H, Olszak K, Przykorska A, Florentz C. Towards understanding human mitochondrial leucine aminoacylation identity. *J Mol Biol*. 2003; 328(5):995–1010. PMID: 12729737
30. Bonnefond L, Giegé R, Rudinger-Thirion J. Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. *Biochimie*. 2005; 87(9–10):873–83. <https://doi.org/10.1016/j.biochi.2005.03.008> PMID: 16164994
31. Wolf YI, Aravind L, Grishin NV, Koonin EV. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res*. 1999; 9(8):689–710. PMID: 10447505
32. Kleeman TA, Wei D, Simpson KL, First EA. Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine. *J Biol Chem*. 1997; 272(22):14420–5. PMID: 9162081
33. Chaudhry SR, Lwin N, Phelan D, Escalante AA, Battistuzzi FU. Comparative analysis of low complexity regions in Plasmodia. *Sci Rep*. 2018; 8(1):335. <https://doi.org/10.1038/s41598-017-18695-y> PMID: 29321589
34. Bonnefond L, Frugier M, Touzé E, Lorber B, Florentz C, Giegé R, et al. Crystal structure of human mitochondrial tyrosyl-tRNA synthetase reveals common and idiosyncratic features. *Structure*. 2007; 15(11):1505–16. <https://doi.org/10.1016/j.str.2007.09.018> PMID: 17997975
35. Yaremchuk A, Krikiviy I, Tukalo M, Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J*. 2002; 21(14):3829–40. <https://doi.org/10.1093/emboj/cdf373> PMID: 12110594
36. Sprinzl M, Vassilenko KS. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res*. 2005; 33(Database issue):D139–40. <https://doi.org/10.1093/nar/gki012> PMID: 15608164
37. Celis JE, Hooper ML, Smith JD. Amino acid acceptor stem of *E. coli* suppressor tRNA<sup>Tyr</sup> is a site of synthetase recognition. *Nat New Biol*. 1973; 244(139):261–4. PMID: 4580701
38. Hou YM, Schimmel P. Modeling with *in vitro* kinetic parameters for the elaboration of transfer RNA identity *in vivo*. *Biochemistry*. 1989; 28(12):4942–7. PMID: 2548595
39. Bedouelle H. Recognition of tRNA(Tyr) by tyrosyl-tRNA synthetase. *Biochimie*. 1990; 72(8):589–98. PMID: 2126463
40. Himeno H, Hasegawa T, Ueda T, Watanabe K, Shimizu M. Conversion of aminoacylation specificity from tRNA<sup>Tyr</sup> to tRNA<sup>Ser</sup> *in vitro*. *Nucleic Acids Res*. 1990; 18(23):6815–9. PMID: 2263446
41. Sherman JM, Rogers K, Rogers MJ, Söll D. Synthetase competition and tRNA context determine the *in vivo* identity of tRNA discriminator mutants. *J Mol Biol*. 1992; 228(4):1055–62. PMID: 1474577

42. Fechter P, Rudinger-Thirion J, Théobald-Dietrich A, Giegé R. Identity of tRNA for yeast tyrosyl-tRNA synthetase: tyrosylation is more sensitive to identity nucleotides than to structural features. *Biochemistry*. 2000; 39(7):1725–33. PMID: [10677221](#)
43. Fechter P, Rudinger-Thirion J, Tukalo M, Giegé R. Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA(Tyr) are conserved but expressed differently. *Eur J Biochem*. 2001; 268(3):761–7. PMID: [11168416](#)
44. Bonnefond L, Frugier M, Giegé R, Rudinger-Thirion J. Human mitochondrial TyrRS disobeys the tyrosine identity rules. *RNA*. 2005; 11(5):558–62. <https://doi.org/10.1261/rna.7246805> PMID: [15840810](#)
45. Messmer M, Blais SP, Balg C, Chenevert R, Grenier L, Lague P, et al. Peculiar inhibition of human mitochondrial aspartyl-tRNA synthetase by adenylate analogs. *Biochimie*. 2009; 91(5):596–603. <https://doi.org/10.1016/j.biochi.2009.02.005> PMID: [19254750](#)
46. Lee CP, Mandal N, Dyson MR, RajBhandary UL. The discriminator base influences tRNA structure at the end of the acceptor stem and possibly its interaction with proteins. *Proc Natl Acad Sci U A*. 1993; 90(15):7149–52.
47. Bour T, Akaddar A, Lorber B, Blais S, Balg C, Candolfi E, et al. Plasmodial aspartyl-tRNA synthetases and peculiarities in *Plasmodium falciparum*. *J Biol Chem*. 2009; 284(28):18893–903. <https://doi.org/10.1074/jbc.M109.015297> PMID: [19443655](#)
48. Hoen R, Novoa EM, Lopez A, Camacho N, Cubells L, Vieira P, et al. Selective inhibition of an apicoplastic aminoacyl-tRNA synthetase from *Plasmodium falciparum*. *ChemBioChem*. 2013; 14(4):499–509. <https://doi.org/10.1002/cbic.201200620> PMID: [23444099](#)
49. Mailu BM, Ramasamay G, Mudeppa DG, Li L, Lindner SE, Peterson MJ, et al. A nondiscriminating glutamyl-tRNA synthetase in the *Plasmodium* apicoplast: the first enzyme in an indirect aminoacylation pathway. *J Biol Chem*. 2013; 288(45):32539–52. <https://doi.org/10.1074/jbc.M113.507467> PMID: [24072705](#)
50. Hu S, Wang M, Cai G, He M. Genetic code-guided protein synthesis and folding in *Escherichia coli*. *J Biol Chem*. 2013; 288(43):30855–61. <https://doi.org/10.1074/jbc.M113.467977> PMID: [24003234](#)
51. Thanaraj TA, Argos P. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci*. 1996; 5(10):1973–83. <https://doi.org/10.1002/pro.5560051003> PMID: [8897597](#)
52. Thanaraj TA, Argos P. Ribosome-mediated translational pause and protein domain organization. *Protein Sci*. 1996; 5(8):1594–612. <https://doi.org/10.1002/pro.5560050814> PMID: [8844849](#)
53. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009; 16(3):274–80. <https://doi.org/10.1038/nsmb.1554> PMID: [19198590](#)
54. Grosjean H, Fiers W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*. 1982; 18(3):199–209. PMID: [6751939](#)
55. Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res*. 2017; gkx1030:<https://doi.org/10.1093/nar/gkx1030>.
56. Vinayak M, Pathak C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci Rep*. 2009; 30:135–48. <https://doi.org/10.1042/BSR20090057> PMID: [19925456](#)
57. Campanacci V, Dubois DY, Becker HD, Kern D, Spinelli S, Valencia C, et al. The *Escherichia coli* YadB gene product reveals a novel aminoacyl-tRNA synthetase like activity. *J Mol Biol*. 2004; 337(2):273–83. <https://doi.org/10.1016/j.jmb.2004.01.027> PMID: [15003446](#)
58. Pütz J, Giegé R, Florentz C. Diversity and similarity in the tRNA world: overall view and case study on malaria-related tRNAs. *FEBS Lett*. 2010; 584(2):350–8. <https://doi.org/10.1016/j.febslet.2009.11.050> PMID: [19931530](#)